RESEARCH REPORT 9
20 JULY 1961
I.E.R. 172-10

A NOTE ON SEMIDEFINITE MATRICES

by

E. Eisenberg

OPERATIONS RESEARCH CENTER

INSTITUTE OF ENGINEERING RESEARCH

UNIVERSITY OF CALIFORNIA - BERKELEY

A NOTE ON SEMIDEFINITE MATRICES

by

Edmund Eisenberg
Operations Research Center
University of California, Berkeley

20 July 1961 Research Report 9

This research has been partially supported by the Office of Naval Research and Bureau of Supplies and Accounts under the Office of Naval Research Contract Nonr-222(83) with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ABSTRACT

It is of general interest to find criteria for a matrix to be positive (or negative)- semidefinite. The usual characterization of semidefinite matrices in terms of their principal minors can be rather laborious to implement practically. We present here an elementary proof of a known alternate characterization of a semidefinite matrix in terms of its null-space and of its largest characteristic value. An iterative procedure is also suggested which may be useful in deciding the semidefiniteness of a matrix.

A NOTE ON SEMIDEFINITE MATRICES

In what follows A will <u>always</u> represent a real, symmetric, $n \times n$ matrix. If, for each $x \in R^n$ (*) it is true that $xAx^T \geq 0$(**) then we say that A is <u>positive-semidefinite</u>, denoted: <u>p.s.d.</u>; if $(xAx^T)(yAy^T) \geq 0$ for all $x, y \in R^n$ we say that A is <u>semidefinite</u>, denoted <u>s.d.</u>. We first prove the following:

<u>THEOREM 1</u>. The following are equivalent:

(i) A is s.d.

(ii) $(xAy^T)^2 \leq (xAx^T)(yAy^T)$, all $x, y \in R^n$

(iii) $x \in R^n$, $xAx^T = 0 \Longrightarrow xA^2x^T = 0$

(iv) $x \in R^n$, $xA^2x^T = 1 \Longrightarrow (xAx^T)^2 > 0$

(v) $x \in R^n$, $xAx^T = 0 \Longrightarrow xA = 0$

<u>PROOF</u>: We show (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (v) $\Longrightarrow$ (i).

<u>(i) $\Longrightarrow$ (ii)</u>

Suppose A is s.d., let $x, y \in R^n$. Consider the real quadratic polynomial p defined by:

$$p(\lambda) = (x + \lambda y)A(x + \lambda y)^T =$$

$$= xAx^T + 2\lambda\, xAy^T + \lambda^2 yAy^T.$$

Since A is s.d., p does not change sign, i.e., its discriminant is non-positive, whence:

$$4(xAy^T)^2 - 4(xAx^T)(yAy^T) \leq 0 ,$$

(*) $R^n = \left\{ x \mid x = (x_1, \ldots, x_n) \text{ and } x_i \text{ is a real number for } i = i, \ldots, n \right\}$.

(**) If $x \in R^n$, x^T denotes the transpose of x.

giving the desired result.

(ii) $\Rightarrow$ (iii).

Suppose $x \in R^n$ and $xAx^T = 0$, then, from (ii), $(xAy^T)^2 \le 0$, i.e., $xAy^T = 0$, for all $y \in R^n$. Thus $xA = 0$, but $xA^2x^T = (xA)(xA)^T = 0$.

(iii) $\Rightarrow$ (iv).

If $x \in R^n$ and $(xAx^T)^2 \le 0$ then $xAx^T = 0$ and, by (iii), $xA^2x^T = 0$, contradicting $xA^2x^T = 1$.

(iv) $\Rightarrow$ (v).

If $x \in R^n$ and $xAx^T = 0$ then, by (iv), $xA^2x^T \le 0$ (because if $xA^2x^T > 0$ then we could normalize x to get $xA^2x^T = 1$, $xAx^T = 0$). However, $xA^2x^T = (xA)(xA)^T$, and thus $xA^2x^T \ge 0$ with equality holding if and only if $xA = 0$.

(v) $\Rightarrow$ (i).

Suppose (i) is false, i.e., there exist $x, y \in R^n$ such that $xAx^T > 0$, $yAy^T < 0$. By suitable normalization [dividing x by $(xAx^T)^{1/2}$ and y by $(-yAy^T)^{1/2}$], we may assume that $xAx^T = 1$, $yAy^T = -1$. Now let:

$$(1) \quad \lambda = -xAy^T + [1 + (xAy^T)^2]^{1/2}$$

$$(2) \quad z = \lambda x + y \quad .$$

We claim that $zA \neq 0$ and $zAz^T = 0$, thus contradicting (v). First, if $zA = 0$ then multiplying (2) by Ax^T and Ay^T we get:

$$0 = \lambda xAx^T + yAx^T = \lambda + xAy^T$$

$$0 = \lambda xAy^T + yAy^T = \lambda xAy^T - 1 \; ,$$

Combining the last two equations:

$$0 = \lambda xAy^T - 1 = (-xAy^T)(xAy^T) - 1 =$$
$$= -1-(xAy^T)^2 \; ,$$

a contradiction, thus $zA \neq 0$. However,

$$\begin{aligned} zAz^T &= (\lambda x + y)A(\lambda x + y)^T = \\ &= \lambda^2 xAx^T + 2\lambda xAy^T + yAy^T \\ &= \lambda^2 + 2\lambda xAy^T - 1 , \end{aligned}$$

and λ was chosen to be precisely one of the two (real) roots of the preceding quadratic polynomial in λ. q.e.d.

Several comments are in order. Obviously, A is s.d. if and only if A is p.s.d. or -A is p.s.d. Condition (ii) of Theorem 1 is a generalization of the Cauchy-Schwartz inequality, namely:

$$(3) \quad (uv^T)^2 \leq (uu^T)(vv^T) \qquad \text{all } u, v \in R^n,$$

for if we take A to be the nxn identity matrix which is clearly p.s.d., we obtain (3) from (ii) - Theorem 1. Condition (v) - Theorem 1, or its obvious equivalents (iii) and (iv), states that if we consider xA, the image under the linear transformation A of a point x in R^n, then A cannot be perpendicular to x unless x is in the null-space of A. Alternately, (v) - Theorem 1 states that if x is <u>not</u> in the null-space of A then its image under A cannot be perpendicular to x.

We proceed next to obtain results which are, in a sense, "refinements" of conditions (ii) (see Lemma 1 below) and (iv) (see Theorem 2) of Theorem 1. Lemma 1 is a generalization of the well known fact, associated with the Cauchy-Schwartz inequality, stating that equality holds in (3) if and only if u, v are linearly dependent. We shall apply Lemma 1 in the proof of Theorem 3.

LEMMA 1

Let A be s.d.. If $x, y \in R^n$ then $(xAy^T)^2 = (xAx^T)(yAy^T)$ if and only if xA, yA are linearly dependent.

PROOF: If, say, $xA = \lambda yA$, where λ is a real number, then $xAy^T = \lambda yAy^T$ while $xAx^T = \lambda yAx^T = \lambda xAy^T = \lambda^2 yAy^T$. Whence it follows that $(xAy^T)^2 = \lambda^2(yAy^T)^2 = (xAx^T)(yAy^T)$.

On the other hand, suppose $(xAy^T)^2 = (xAx^T)(yAy^T)$. If $xAx^T = 0$ or $yAy^T = 0$ then, by (v) - Theorem 1, $xA = 0$ or $yA = 0$ and we certainly can conclude that xA, yA are linearly dependent. Otherwise, say, $xAx^T > 0$ and $yAy^T > 0$, consequently $xAy^T \neq 0$. Let $\rho = \text{signum}(xAy^T)$ and let:

$$\alpha = (yAy^T)^{1/2}$$

$$\beta = -\rho(xAx^T)^{1/2},$$

then $\alpha, \beta \neq 0$ and:

$$(\alpha x + \beta y)A(\alpha x + \beta y)^T = \alpha^2 xAx^T + \beta^2 yAy^T + 2\alpha\beta xAy^T =$$

$$= 2(xAx^T)(yAy^T) - 2\rho(xAy^T)(xAx^T)^{1/2}(yAy^T)^{1/2} =$$

$$= 2(xAx^T)(yAy^T) - 2|xAy^T|(xAx^T)^{1/2}(yAy^T)^{1/2} =$$

$$= 2(xAx^T)(yAy^T) - 2(xAx^T)(yAy^T) = 0 .$$

Thus, $(\alpha x + \beta y)A(\alpha x + \beta y)^T = 0$ and, by (v) - Theorem 1, $0 = (\alpha x + \beta y)A = \alpha xA + \beta yA$. q.e.d.

The preceding lemma was motivated, in part, by an examination of (ii) - Theorem 1 in case A is the identity matrix, in that case (since the square of the identity is the identity), (iv) - Theorem 1 states: $x \in R^n$, $xx^T = 1$ implies $(xx^T)^2 > 0$, which is, of course, true. We notice, though, that $(xAx^T)^2$ has then a <u>positive</u> lower bound, namely 1. In general, this will be the case, i.e., a positive lower bound will exist for $(xAx^T)^2$ in (iv) - Theorem 1, whenever A is s.d.. Clearly, when A is identically zero any positive number will serve as a lower bound, because there is no $x \in R^n$ for which $xA^2x^T = 1$, thus we will exclude $A = 0$ in the next theorem:

THEOREM 2

Suppose A is p.s.d. and $A \neq 0$, then there exist a positive real number μ and an $x_0 \in R^n$ such that:

(4) $\quad x \in R^n,\ xA^2x^T = 1 \Rightarrow xAx^T \geq \mu$

(5) $\quad x_0A^2x_0^T = 1$ and $x_0Ax_0^T = \mu$.

PROOF: Let

$$X = \left\{ x \mid x \in R^n \text{ and } xA^2x^T = 1 \right\}$$

$$\mu = \inf_{x \in X} xAx^T .$$

Since A is p.s.d. and $A \neq 0$, μ is well defined and in fact $\mu \geq 0$ and satisfies (4). By definition of μ, there exists a sequence x_k such that

(6) $\quad x_k \in X \quad$ for $k = 1, 2, \ldots$

(7) $\quad x_kAx_k^T$ converges to μ.

We consider two cases:

Case 1. The sequence x_k has a bounded subsequence. In this eventuality the x_k have a point of accumulation x_0, for which it must be true (by (6) and (7) and because X is closed) that $x_0 \in X$ and $x_0Ax_0^T = \mu$. Thus x_0 satisfies (5). That μ is positive then follows from (v) - Theorem 1. The two preceding facts, together with the remark above that μ satisfies 4, complete the proof.

Case 2. The sequence $\{x_k\}$ has no bounded subsequence, i.e., we may assume that $|x_k| = (x_k x_k^T)^{1/2} \to \infty$, and $|x_k| > 0$, $k = 1, 2, \ldots$. We define another sequence $\{y_k\}$ by:

$$y_k = \frac{x_k}{|x_k|} .$$

Now, $y_k A y_k^T$ converge to zero, because $x_k A x_k^T$ converge to μ and also $y_k A^2 y_k^T$ converge to zero, because $x_k A^2 x_k = 1$ all k. However, $|y_k| = 1$, thus the y_k's have an accumulation point y, for which it must be true that $yAy^T = 0$. Thus $yA = 0$ by (v) - Theorem 1.

Next we observe that from the definition of y and the y_k's it follows that whenever y has a non-zero component then infinitely many x_k's have the same component non-zero, and in fact of the same sign. We may assume that an appropriate subsequence of x_k has been selected so that whenever y has a positive (negative) component then <u>all</u> the x_k's have the same component positive (negative). Now, if $\{\lambda_k\}$ is any sequence of real numbers then:

$$(x_k + \lambda_k y)A(x_k + \lambda_k y)^T = x_k A x_k^T$$

and $$(x_k + \lambda_k y)A^2(x_k + \lambda_k y)^T = x_k A^2 x_k^T ,$$

because $yA = 0$. We can thus replace x_k by $x_k + \lambda_k y$, $k = 1, 2, \ldots,$ and (6) and (7) will still hold. However, by an appropriate choice of λ_k we can reduce the number of non-zero components in each of the x_k's, eventually (repeating the above process, if necessary) we obtain a sequence $\{x_k\}$, satisfying (6)-(7) and which has an accumulation point, thus reducing it to case 1. q.e.d.

As an immediate consequence of Theorem 2 we can "strengthen" (iv) - Theorem 1.

Corollary

If A is s.d. and $A \neq 0$ then

$$\text{minimum} \left\{ \left(xAx^T\right)^2 \,\middle|\, x \in R^n \text{ and } xA^2x^T = 1 \right\}$$

exists and is positive.

PROOF: As noted before, if A is s.d., then either A is p.s.d. or $-A$ is p.s.d., in either case the square of the μ in Theorem 2 is the required minimum and the x_0 of the same theorem is the required minimizing x.

The μ and x_0 of Theorem 2 are, as one might expect, intimately related to the characteristic values of A. This is brought forth in the next theorem.

THEOREM 3

Let A be p.s.d., $A \neq 0$. Let μ and x_0 be as in Theorem 2 and let λ_n be the largest characteristic value of A, then $\lambda_n = \mu^{-1}$ and x_0A is a characteristic vector of A corresponding to λ_n.

PROOF: Suppose λ is any characteristic value of A, i.e., there exists an $x \in R^n$, $x \neq 0$, such that $xA = \lambda x$, whence $xA^2x^T = \lambda xAx^T$. If $\lambda = 0$ then certainly $\lambda \leq \mu^{-1}$. Assuming $\lambda \neq 0$, it follows that $xA \neq 0$ (because $x \neq 0$) and thus, by (v) - Theorem 1, $xAx^T > 0$. Let $y = (xA^2x^T)^{-1/2}x$, then $yA^2y = 1$ and, by definition of μ, $yAy^T \geq \mu$. However $yAy^T = (xA^2x^T)^{-1}(xAx^T) = = \lambda^{-1}$, thus $\lambda \leq \mu^{-1}$. We have just demonstrated that $\lambda \leq \mu^{-1}$ for any characteristic value λ of A, thus $\lambda_n \leq \mu^{-1}$.

To complete the proof of this theorem it will suffice to show that there is a characteristic value λ of A such that $\lambda = \mu^{-1}$, and $(x_0A)A = \lambda(x_0A)$, x_0 being as in Theorem 2. Let $x = x_0$ be a minimizing x_0 in question. Since A and A^2 are p. s. d. (the square of any real symmetric matrix is p. s. d.), and $xA \neq 0$ ($xAx^T = x_0Ax_0^T = \mu > 0$), it follows that $xA^3x^T = (xA)A(xA)^T > 0$, and $xA^4x = (xA)A^2(xA)^T > 0$. Thus, if we define

$$(8) \qquad \rho = 2(xA^3x^T)(xA^4x^T)^{-1}$$

then ρ is positive. Next let

$$(9) \qquad y = x - \rho x A \; .$$

The motivation for the above definition of y is as follows: we know x minimizes a certain function, namely xAx, since we wish to derive from this fact some properties of x we examine how xAx will change in the direction of its gradient, namely $2xA$. As defined in (9), y is a translation from x precisely in the direction of that gradient, the particular value of ρ chosen is designed to keep y within the "feasibility" set, i.e., $yA^2y = 1$. We check next the last mentioned condition:

$$\begin{aligned} yA^2y^T &= (x - \rho xA)A^2(x - \rho xA)^T = \\ &= xA^2x^T - 2\rho xA^3x^T + \rho^2 xA^4x^T = \\ &= 1 - 2\rho\left[xA^3x^T - \frac{\rho}{2}(xA^4x^T)\right] \\ &= 1 - 2\rho\left[xA^3x^T - (xA^3x^T)(xA^4x^T)^{-1}(xA^4x^T)\right] \\ &= 1 \; . \end{aligned}$$

One can, incidentally, readily check that the particular value of ρ, as given in (8), is the only value of ρ (other than $\rho = 0$) which yields $yA^2y = 1$. Now, since $yA^2y^T = 1$, we must have, by definition of μ,

(10) $yAy^T - xAx^T \geq 0$.

However,

$$yAy^T - xAx^T = (x-\rho xA)A(x-\rho xA)^T - xAx^T =$$
$$= -2\rho xA^2x^T + \rho^2 xA^3x^T =$$
$$= 2\rho\left[\frac{\rho}{2}(xA^3x^T) - (xA^2x^T)\right] =$$
$$= 2\rho(xA^4x^T)^{-1}\left[(xA^3x^T)^2 - (xA^2x^T)(xA^4x^T)\right] .$$

Thus, since $\rho > 0$, $(xA^4x^T)^{-1} > 0$ and because (10) holds, we have:

(11) $(xA^3x^T)^2 \geq (xA^2x^T)(xA^4x^T)$.

We now refer to inequality (3), which is a special case of (ii) - Theorem 1 with A being the identity, letting $u = xA$, $v = xA^2$ we get:

(12) $(xA^3x^T)^2 \leq (xA^2x^T)(xA^4x^T)$.

Combining (11) and (12), we get:

(13) $(xA^3x^T)^2 = (xA^2x^T)(xA^4x^T)$.

However, from Lemma 1, again with A being the identity matrix, we then know that xA, xA^2 are linearly dependent. Since $xA \neq 0$, it follows that there is a real number λ such that $xA^2 = \lambda xA$, multiplying by x^T : $1 = xA^2x^T = \lambda xAx^T$, and $\lambda = (xAx^T)^{-1} = \mu^{-1}$. q.e.d.

As a final general result, we specialize (ii) - Theorem 1, and Lemma 1, for the case where A is non-singular.

THEOREM 4

Let A be p. s. d. and non-singular then,

$$(14) \quad (uv^T)^2 \leq (uAu^T)(vA^{-1}v^T) \quad \text{all } u, v \in R^n$$

and equality holds above if and only if u, vA^{-1} are dependent.

PROOF: We first note that A^{-1} must be symmetric because $AA^{-1} = I$, thus $I^T = I = (AA^{-1})^T = (A^{-1})^T A^T = (A^{-1})^T A$. But the inverse is unique, thus $A^{-1} = (A^{-1})^T$. Next, let $u, v \in R^n$, we let

$$(15) \quad x = u, \; y = vA^{-1} .$$

One readily checks that:

$$xAy^T = uv^T, \; xAx^T = uAu^T, \; yAy^T = vA^{-1}v^T .$$

Thus the desired inequality (14) follows from (ii) - Theorem 1. Now if (14) is actually an equation, then from Lemma 1, using x, y as defined in (15), we get u, vA^{-1} are linearly dependent. The converse also follows readily.

q. e. d.

Note: The condition of equality in (14) is directly connected with characteristic vectors of A (and of course, those of A^{-1}), for suppose (14) is an equation and $u = v \neq 0$, then one sees immediately that $uA = \lambda u$ for some real number λ. The corresponding converse also holds in this case.

An iterative scheme, for deciding the definiteness of A, based on the proof of Theorem 3 might go as follows:

(a) By examining the diagonal elements of A we have decided that, if at all, A is p. s. d.

(b) We have an x such that $xA \neq 0$; if $xA^2x \leq 0$ then A is not p.s.d., if $xA^2x > 0$ normalize x so that $xA^2x = 1$ and proceed to (c)

(c) We have an x such that $xA \neq 0$, $xA^2x = 1$; perform the transformation given by (8) and (9). There are three cases:

<u>Case 1.</u> if $yAy^T > xAx^T$ then A is not p.s.d.

<u>Case 2.</u> if $yAy^T < xAx^T$ return to beginning of (c), using y as the new "test" vector.

<u>Case 3.</u> if $yAy^T = xAx^T$ we have isolated a characteristic vector of A, return to (b) using, as x, a vector independent of all characteristic vectors thus far obtained.

The preceding is, of course, "informal" in the sense that the iterative procedure described above has not been shown to coverge.

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

Head, Logistics and Mathematical
Statistics Branch
Office of Naval Research
Washington 25, D. C.

C.O., ONR Branch Office
Navy No. 100 F.P.O.
New York City, New York

ASTIA Document Service Center
Arlington Hall Station
Arlington 12, Virginia

Office of Technical Services
Department of Commerce
Washington 25, D. C.

Technical Information Officer
Naval Research Laboratory
Washington 25, D. C.

C.O., ONR Branch Office
346 Broadway, New York 13, N.Y.
Attn: J. Laderman

C.O., ONR Branch Office
1030 East Green Street
Pasadena 1, California
Attn: Dr. A. R. Laufer

Professor Russell Ackoff
Operations Research Group
Case Institute of Technology
Cleveland 6, Ohio

Professor Kenneth J. Arrow
Serra House
Stanford University
Stanford, California

Professor G. L. Bach
Carnegie Institute of Tech.
Planning and Control of
Industrial Operations
Schenley Park
Pittsburgh 13, Penn.

Professor A. Charnes
The Technological Institute
Northwestern University
Evanston, Illinois

Professor L. W. Cohen
Math. Dept., Univ. of Maryland
College Park, Maryland

Professor Donald Eckman
Director, Systems Research
Center, Case Inst. of Tech.
Cleveland, Ohio

Professor Lawrence E. Fouraker
Dept. of Economics, The
Pennsylvania State University
State College, Pennsylvania

Professor David Gale
Dept. of Math., Brown University
Providence 12, Rhode Island

Professor L. Hurwicz
School of Business Administration
University of Minnesota
Minneapolis 14, Minnesota

Professor James R. Jackson
Management Sciences Research
Project, Univ. of California
Los Angeles 24, California

Professor Samuel Karlin
Dept. of Math., Stanford Univ.
Stanford, California

Professor C. E. Lemke
Dept. of Mathematics
Rensselaer Polytechnic Institute
Troy, New York

Professor W. H. Marlow
Logistics Research Project
The Geo. Wash. University
707 - 22nd Street, N.W.
Washington 7, D. C.

Professor Oskar Morgenstern
Economics Research Project
Princeton University
92 A Nassau Street
Princeton, New Jersey

Professor R. Radner
Department of Economics
University of California
Berkeley, California

Professor Stanley Reiter
Department of Economics
Purdue University
Lafayette, Indiana

Mr. J.R. Simpson, Bureau
of Supplies and Accounts
(Code W31) Navy Department
Washington 25, D. C.

Professor A. W. Tucker
Dept. of Mathematics
Princeton University
Princeton, New Jersey

Professor J. Wolfowitz
Dept. of Mathematics
Lincoln Hall, Cornell Univ.
Ithaca 1, New York